Naval Research Laboratory

Washington, DC 20375-5000

AD-A227 331

DTIC FILE COPY

Movable-Boundary Channel-Access Schemes for Integrated Voice/Data Networks
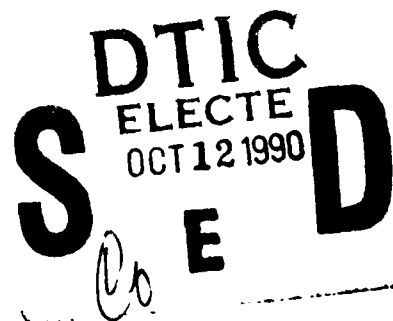
J. E. WIESELTHIER AND A. EPHREMIDES*

Communication Systems Branch
Information Technology Division

*Electrical Engineering Department
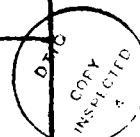*University of Maryland*
College Park, Maryland 20742

and

LOCUS, Inc.
Alexandria, Virginia 22303

August 20, 1990

DTIC
ELECTE
OCT 12 1990
S   E   D

90 10 1

REPORT DOCUMENTATION PAGE

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE 1990 August 20	3. REPORT TYPE AND DATES COVERED Interim 1/89 - 6/89

4. TITLE AND SUBTITLE
Movable-Boundary Channel-Access Schemes
for Integrated Voice/Data Networks

5. FUNDING NUMBERS
PE: 61153N
PR: RR021-0542
WU: DN159-036

6. AUTHOR(S)
Jeffrey E. Wieselthier
and Anthony Ephremides

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Naval Research Laboratory
Washington, DC 20375-5000

8. PERFORMING ORGANIZATION REPORT NUMBER

NRL Memorandum
Report 6706

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Office of Naval Research
Arlington, VA 22217

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

* Anthony Ephremides is with the University of Maryland and LOCUS, Inc.

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

Channel-access protocols for integrated radio networks must reflect the different communication requirements associated with voice and data, as well as the impact each type of traffic has on the other. In this report we introduce a new class of protocols that we have developed for this application, which are known as the Voice/Data Interleaved-Frame Fixed-Length (VD-IFFL) protocols. As a first step in the discussion of the VD-IFFL protocols, we introduce the IFFL protocols, which are similar to the Interleaved-Frame Flush-Out (IFFO) protocols (a class of schemes developed for data-only applications) except that they have a fixed frame length, a desirable feature in voice communication. The VD-IFFL protocols use a movable-boundary mechanism to share the channel between voice and data traffic. Voice traffic is handled on a reservation basis, while data traffic is handled using a hybrid IFFL scheme that combines reservation and contention. These protocols are characterized by infinite Markov chains, whose transition probabilities have been evaluated exactly.

14. SUBJECT TERMS
Multiple access
Communications network
Voice/data integration
Markov chain

15. NUMBER OF PAGES
48

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

NSN 7540-01-280-5500

CONTENTS

MOVABLE-BOUNDARY CHANNEL-ACCESS SCHEMES FOR INTEGRATED VOICE/DATA NETWORKS

1. INTRODUCTION

The design of channel-access protocols for any specific application should reflect the characteristics of the communication traffic that must be supported by the network as well as the nature of the communication medium. In particular, when considering integrated voice/data radio networks, channel access methods must be developed that reflect the different requirements on delay and error rate that are associated with voice and data traffic as well as the impact each type of traffic has on the other.

A variety of approaches are available for channel access in data networks. Descriptions of many of these schemes may be found in survey articles by Tobagi [1] and Lam [2]. Depending on the nature of the traffic, either contention-based or contention-free schemes or their hybrids can be used. Many comparative discussions of channel access protocols are found in the literature, and we do not intend to discuss their merits here. However, the problem of channel access in integrated radio networks has not received much attention.

In this report we first review briefly some of the major issues associated with channel access in integrated radio networks. A more complete presentation is given in [3]. We then discuss some approaches found in the literature. The main purpose of this report, however, is the introduction and analysis of a new protocol for integrated voice/data communication, primarily in satellite networks. This protocol is a modification of the Interleaved-Frame Flush-Out (IFFO) protocols for data traffic, which were introduced by Wieselthier and Ephremides [4, 5] a decade ago. We briefly review the IFFO protocols and the mathematical model used to describe them. These protocols are characterized by a frame length that adapts to bursty channel traffic, resulting in

extremely high efficiency.

We then consider variations of the IFFO protocols (still considering only data traffic) in which the frame length is kept constant. These are known as the Interleaved-Frame Fixed Length (IFFL) and Non-Interleaved-Frame Fixed-Length (NIFFL) schemes. The property of constant frame length is desirable for voice traffic, which is generally characterized by the need for near-real-time delivery but, more importantly, with constant delay. Using this framework, we then extend the IFFL protocols for operation in integrated networks by incorporating a movable-boundary mechanism to share the channel between voice and data traffic; the new protocols are called the Voice/Data IFFL (VD-IFFL) protocols.

We present the transition probabilities of the underlying Markov chains for all of these protocols. In all cases these transition probabilities are determined exactly; however, since the chains are infinite, truncation is needed for numerical evaluation. Like IFFO, the IFFL schemes are characterized by first-order Markov chains. In contrast, the NIFFL schemes are characterized by second-order Markov chains, and the VD-IFFL schemes are characterized by a two-dimensional Markov chain. In these cases we exploit special features of the system to reduce complexity, without requiring any approximations, thereby making numerical solution possible. For the case of VD-IFFL we have developed an approximation to simplify the system description. Since an exact description is also available, it will be possible to verify the accuracy of this approximation in the future. At this point, numerical results are available only for the IFFO protocols. However, the analytical framework presented here permits an evaluation of the other protocols as well.

Other framed data protocols can also be modified for integrated voice/data operation in a similar manner. For example, we briefly discuss the use of the Framed ALOHA protocol [6] in an integrated network. However, in this report we concentrate on the IFFO protocols and their extensions.

2. CHARACTERISTICS OF VOICE TRAFFIC

Voice traffic is characterized by the need for delivery as a continuous stream in near-real time. More importantly, the delay must be nearly constant throughout the duration of each talkspurt. It is assumed that buffering is not possible.* Calls are blocked if channel resources are not immediately available; acceptance of a voice call requires a continuous commitment of channel resource (e.g., a time slot in every frame or a fixed portion of the bandwidth for frequency division systems) for the entire duration of the call. Thus, voice traffic is quite different from data traffic, which may be either bursty or regular in nature, and which may consist of either one or several packets. In either case, data is characterized by a need for very low packet-error probability but not for real-time delivery. Delay requirements depend on the nature of the traffic and may be different for different classes of traffic in the network. Buffering of data packets is permitted.

The need to support the requirements of voice traffic results in the need for contention-free channel access once a call has been set up. Reservation schemes, which can maintain throughput levels near channel capacity, are the logical choice for voice calls. Many reservation schemes are proposed in the literature, and most of them are modifications of the demand-assignment scheme originally proposed by Roberts [7]. Once a reservation for a voice call is successfully made and acknowledged, the user is allowed contention-free access to the channel until the end of the call. This is signaled by the user's end-of-message indicator, at which point the channel (time slot or frequency slot) becomes available for assignment to another user.

---

* Even if voice buffering is permitted, the requirements for continuity and constant delay do not change.

3. BOUNDARY SCHEMES FOR VOICE/DATA MULTIPLEXING

The goal of voice/data integration is to share network resources efficiently between these two classes of traffic while satisfying the performance requirements of both. Networking schemes are needed that can simultaneously provide the equivalent of circuit switching for voice and packet switching for data. Most studies of voice/data integration have addressed this problem from the perspective of operation at a single node, with the goal of developing a multiplexing scheme that satisfies the real-time delivery requirements of a sufficiently large number of voice calls, while providing for adequate data throughput with acceptable delays. Our discussion focuses on the "movable-boundary" scheme, which has emerged as the prime hybrid system of integrated switching. The term slotted envelope network (SENET) has been often applied to describe any boundary scheme for voice/data integration. However, we prefer to use the more general term movable boundary to describe this type of multiplexing mechanism. A thorough discussion of movable-boundary schemes for multiplexing is presented in [3]. It is straightforward to extend the boundary concept from the realm of multiplexing to that of channel access, although few results are available in the literature.

The boundary scheme is based on a TDMA frame structure (Fig. 1). The fixed-length TDMA frame is partitioned into two compartments, with voice being circuit-switched in one and data being packet-switched in the other. The boundary between the two compartments can be either fixed or movable. In the fixed-boundary scheme, voice is transmitted only in the slots of one compartment; any unused slots from the data compartment are left unused. Data are handled in a similar way. Under the movable-boundary scheme, data traffic is allowed to use any idle slots of the voice compartment, resulting in higher bandwidth utilization. However, voice traffic is not permitted to use unused data slots.

The acceptance of a voice call by the communication system implies a long-term commitment of a channel (in this case a TDMA slot) to support the call. It is generally

assumed that a voice call cannot be interrupted once it is assigned a channel. Data traffic requires only a short-term commitment, i.e., ι .e packet at a time.* Since each data packet occupies only one slot at a time, data traffic does not interfere with voice traffic: the voice slot that was borrowed for data reverts to its original status as a voice slot immediately when needed for this purpose.

The use of movable-boundary schemes permits the use of dynamic optimization techniques that adapt to channel traffic. Based on traffic, the position of the boundary can be chosen to optimize system performance. Allocation of too small a fraction of the channel to voice traffic can result in the blockage of too many calls (voice traffic cannot use data slots, even when they are not in use); allocation of too large a fraction to voice can result in excessive queueing delay for data (or packet loss if buffer capacities are exceeded). In general, a performance measure can be defined as a weighted sum of voice and data criteria. Full and accurate analysis of movable-boundary schemes is very difficult, except for the simplest of examples.

The movable-boundary scheme for multiplexing at a single node has been studied extensively since it was first proposed by Kummerle [8] and Zafiropoulo [9]. Shortly thereafter, Coviello and Vena [10] described in detail the operation of a movable-boundary scheme developed for a T1 carrier, which they named the Slotted Envelope Network (SENET). A summary of the most important papers in this area is presented in [3]. The boundary method has also been applied to the channel access problem, as we now discuss.

4. EARLIER STUDIES OF CHANNEL ACCESS IN INTEGRATED NETWORKS

The problem of channel access in integrated radio networks has not received much attention in the literature. Recently, Suda et al. [11] and Wu and Li [12] studied

---

* Although data messages can consist of more than one packet, each packet of a multipacket message can be treated separately by the network since there is no need for uniformity of delay.

protocols for access to satellite channels. These schemes are characterized by a fixed-length frame structure that contains reservation channels and information channels (some of which are allocated for voice and the remainder for data). Voice is handled on a reservation basis in both of these studies. Once a reservation for a voice call is made successfully, one slot per frame is allocated to the call until its completion. There are a number of differences in the two models, however, particularly in the channel access mechanism for data. Nevertheless, these schemes bear a strong similarity to the ideas developed in [4] for the IFFO protocols, which inspired the introduction of the new protocols that form the main contribution of this report.

Suda et al. considered only the use of the slotted ALOHA random-access protocol for data traffic (making the assumption of infinite buffer capacity) and considered operation only under a fixed-boundary scheme. They noted that extensions of their analysis to movable-boundary schemes and finite buffer models would be difficult. Wu and Li considered three options for data traffic, namely, random access, reservation, and hybrid (i.e., combining features of random access and reservation). Where reservations are used (i.e., for all voice schemes and for random access and hybrid data schemes) a distributed reservation scheme was used. Both fixed- and movable- boundary schemes were considered for sharing the channel resource between voice and data traffic. However, their analysis was based on an overly simplified approximation.

Spread-Spectrum Considerations

The use of spread-spectrum signaling, necessitated by antijam considerations in many military applications, leads naturally to the use of code-division multiple-access (CDMA) techniques. The use of CDMA provides an environment that is radically different from that of narrowband time-domain operation. A great deal of flexibility in channel-access protocol design can be achieved by taking advantage of the multiple-user capability, the selective-addressing capability, and the selective-reception capability of CDMA signaling [13]. In the future, we will develop schemes for channel access in

6

integrated networks that exploit the ability of the CDMA channel to support several transmissions (on quasiorthogonal frequency-hoppping patterns, or codes) simultaneously. Issues associated with CDMA signaling in integrated networks are discussed in more detail in [3]. Thus far, only Soroushnejad and Geraniotis [14, 15, 16] have published results in this area. The present report, however, does not address CDMA considerations.

5. THE COMMUNICATION SYSTEM: DATA-ONLY OPERATION

In this section we describe the communication system for data-only operation. It is then straightforward to extend the model to integrated networks, as discussed in Section 9.

We consider M ground-based users (terminals) that communicate among themselves via a transponder, which broadcasts all messages it receives to all members of the user population. A method is needed to determine how the users should schedule their packet transmissions to avoid destructive "collisions," which occur when two or more terminals transmit simultaneously, and to maintain high efficiency. This is known as the channel-access problem.

In much of this report we assume that the transponder is located on a geosynchronous satellite, which results in a substantial round-trip propagation time of approximately 0.27 second. Since we consider fixed-length packets as units of transmission (each requires one time "slot" for transmission), this delay may be expressed in terms of packet length as R slots. The value of R is an important system parameter that has a great impact on protocol operation and performance. Typically, R is of the order of 10 - 12 slots for satellite-based systems.* However, for ground-based transmissions R can be significantly smaller (much less than one slot duration), resulting

* The value of R = 12 used in [4] was based on a data rate of 50,000 bits/s and a packet length of 1125 bits. This data rate and packet length are typical of multiple access studies that have appeared in the literature.

7

in simplified analysis and improved performance. In this report we emphasize satellite-based systems because they are more challenging to model, but we also indicate how the analysis would be modified for ground-radio systems.

It is assumed that each user has an infinite buffer in which it stores the arriving packets, which are assumed to form a Bernoulli process with rate λ in every slot. The total arival rate is, therefore, $M\lambda$ packets per slot, which is equal to the throughput rate under stable operation since no packets are rejected.

6. A MARKOV CHAIN MODEL FOR THE IFFO PROTOCOLS

The basic structure for the IFFO protocols, as shown in Fig. 2, is a reservation structure, in which the unreserved slots may be used for transmission on a contention basis. The first slot of each frame, which consists of M "minislots" that are exclusively allocated to the M terminals in (contention-free) TDMA fashion, is known as the status slot; it is used by each of the terminals to reserve a transmission slot for each of the packets that were generated in the previous frame.*

It is assumed that all reservation minipackets are received successfully by all terminals following the round-trip propagation delay of R slots. We define:

R_k = total number of reserved slots in frame k;

N_k = total number of contention (unreserved) slots in frame k;

L_k = total length (in slots) of frame k.

Thus the status slot is followed by R_k reserved slots. It is required that each frame have a length of at least R slots to ensure that the reservation information generated at the

---

* Clearly, the number of terminals that can be accommodated by this protocol is limited by the number of minislots that can be established in one slot duration. These minipackets could be quite short because the only information they would have to deliver is the number of packets that arrived at the terminal during the previous frame. Alternatively, a contention-based channel access scheme could be used for the reservation minipackets, but such a system has not yet been analyzed. One might also consider the use of a CDMA reservation channel in applications where the satellite can monitor several codes simultaneously.

beginning of the frame is received before the start of the next frame. If $R_k < R-1$, the remaining $N_k = R - 1 - R_k$ slots in the frame are used for transmission on a contention basis, as in the so-called slotted-ALOHA protocol.* If $R_k > R-1$, additional slots are added to accommodate all of the reservations. Therefore,

$$N_k = \max(R-1-R_k, 0),$$

and

$$L_k = \max(R_k+1, R) = 1 + R_k + N_k.$$

Since the frame length expands to accommodate all packets for which reservations have been received, and since there is one slot of overhead per frame regardless of frame length, throughput rates arbitarily close to one packet per slot can be realized. However, packet delay increases rapidly as throughput approaches one.

The quantity of interest that needs to be tracked is R_k, which evolves as a first-order Markov chain. A complete discussion of the dynamics is presented in [4]. Here we summarize the derivation and show the resulting transition probabilities.

The operation of the IFFO protocols is illustrated in Fig. 3. Each packet arriving in frame k is known as a *k-packet*. Reservation minipackets for all k-packets will be transmitted in the first slot of frame $k+1$. Although reservations are transmitted for all k-packets, some of these reservations may not be needed, owing to the possibility of successful transmission in contention slots. It is easy to cancel the unneeded reservations because knowledge of the outcome of all contention transmissions in frame k will be available to all users not later than the beginning of frame $k+2$.

Let

A_k = total number of packet arrivals in frame k, summed over all terminals;

---

* Time slots are not assigned to specific users under contention-based operation. Whenever two or more users transmit in the same slot, all packets involved in the collision are assumed to be destroyed. They will be retransmitted in reserved slots, as we soon discuss.

S_k = number of successful contention transmissions in frame k.

Then,

$$R_{k+2} = A_k - S_k.$$

Note that R_{k+2} is totally independent of A_{k+1}, L_{k+1} and R_{k+1}. It depends only on R_k and on what happens during frame k. Thus we can split the process $\{R_k\}$ into two interleaved Markov chains, which may be denoted $\{R_{2j}\}$ and $\{R_{2j+1}\}$. The reserved slot process in even-numbered frames is independent of the reserved slot process in odd-numbered frames. These processes have identical statistics and may be analyzed separately. In Section 8 we consider modifications of the IFFO protocols under which the even- and odd-numbered slots are no longer independent. In that case a second-order Markov chain is needed to model system behavior, resulting in a system description that is considerably more complicated.

There are several versions of the IFFO protocols, each of which is characterized by a different transmission procedure in the unreserved slots:

1) Pure Reservation IFFO (PR-IFFO): The unreserved slots are not used for contention; they simply remain idle and wasted. All packets that arrive in frame k are transmitted in the reserved slots of frame $k+2$.

2) Fixed Contention IFFO (F-IFFO): The transmission policy depends on the slot number in which packets arrive.

a) A packet arriving in slot n, for $n \in (R_k+1, R-1)$, will be transmitted in slot $n+1$, i.e., in each contention slot, each terminal will transmit the packet that may have arrived in the previous slot. Each colliding packet will be retransmitted in a reserved slot in frame $k+2$.

b) All packets arriving during the first R_k slots of frame k will not be allowed to contend because of the high risk of collision, and will be assigned reserved slots for transmision during frame $k+2$.

10

c) All packets arriving during slot R (i.e., the last slot in the frame) will be assigned reserved slots in frame $k+2$.

3) Controlled Contention IFFO (C-IFFO): In this version, the transmission procedure is state-dependent. A complete description and approximate analysis of this protocol are presented in [4].

The Transition Probability Matrices for IFFO

For PR-IFFO, since $R_{k+2} = A_k$, it is easy to see that the elements of the transition probability matrix for R_k can be written as

$$p_{ij} \triangleq Pr(R_{k+2} = j | R_k = i) = \begin{cases} \binom{MR}{j} \lambda^j (1-\lambda)^{MR-j}, & 0 \leq i \leq R-1 \\ \binom{M(i+1)}{j} \lambda^j (1-\lambda)^{M(i+1)-j}, & i \geq R-1. \end{cases}$$

The evaluation of the combinatorial expressions in the above equation raises an interesting computational problem. For large values of M (values up to 50 have been considered) and i (as great as 700) the values of the combinatorial expressions become extremely large, while the values of the exponential expressions become extremely small. To avoid the resulting computational problems, we recognize that, since the arrival process in each slot is an independent sequence, the probability mass function (pmf) of the number of arrivals in $i+1$ slots is simply the convolution of $i+1$ pmf' of the number of arrivals in a single slot. Thus we can perform the computation by convolving expressions that have less extreme values. We have

$$a(j) = Pr(j \text{ arrivals in one slot}) = \binom{M}{j} \lambda^j (1-\lambda)^{M-j}.$$

Therefore, for $i \geq R-1$,

$$p_{ij} = a(j) * a(j) * \cdots * a(j),$$

where * represents the convolution operator and the expression is the convolution of $i+1$ pmf's of the form $a(j)$. Similarly, for $i < R-1$, p_{ij} is the convolution of R $a(j)$'s. All

11

convolutions are performed numerically.

For F-IFFO we can easily show that (see [4])

$$p_{ij} \triangleq Pr(R_{k+2} = j | R_k = i) = \binom{M(i+1)}{j} \lambda^j (1-\lambda)^{M(i+1)-j} * c_{N_k}(j),$$

where

$$c_{N_k}(j) = c(j) * c(j) * \cdots * c(j)$$

is the convolution of N_k (which depends on $R_k = i$) pmf's of the form $c(j)$, which is the probability that j packets are transmitted unsuccessfully in a contention slot. In particular,

$$c(0) = (1-\lambda)^{M-1}[(1-\lambda) + M\lambda]$$

$$c(1) = 0$$

$$c(j) = \binom{M}{j} \lambda^j (1-\lambda)^{M-j}, \quad 2 \leq j \leq M.$$

Thus $c_{N_k}(j)$ is the pmf of the number of unsuccessful (colliding) packets in a frame with N_k contention slots. The expression to the left of the convolution operator in the transition probability for F-IFFO is the pmf of the number of arrivals in frame k that do not attempt transmission in contention slots (because of the slot number in which they arrived; see the above discussion on transmission policy). Whenever $R_k \geq R-1$, the convolution vanishes because $N_k = 0$; thus the p_{ij} are the same as those for PR-IFFO. Therefore, the performance of F-IFFO is closely bounded by that of PR-IFFO in the limit of high input rates.

On Obtaining Equilibrium Results for the IFFO Protocols

The equilibrium pmf for the number of reserved slots per frame is needed to evaluate the two steady-state performance indexes that have been considered for the IFFO protocols, i.e., the expected time spent in the system per packet, and the expected number of reserved slots per frame. To evaluate performance we need to compute the

stationary vector of the chain R_k. Expressions for expected packet delay are presented in [5] and [17].

An equilibrium pmf for R_k does, in fact, exist under the IFFO protocols, as long as the input rate is less than one packet per slot. In this case, R_k, which is irreducible and aperiodic, is easily shown to be ergodic by Foster's theorem [18].

Note that the chain that describes R_k has an unbounded (hence, infinite) number of states. To obtain a numerical solution, we truncate the probability vector and transition probability matrix to some finite dimension N. N can be chosen sufficiently large so that the effect of truncation error is small. An $N \times N$ transition probability matrix permits the modeling of up to $N-1$ reserved slots per frame (since the N elements in the probability vector range from 0 to $N-1$). The truncation operation involves setting $Pr(j = N-1|i) = 1 - Pr(j \leq N-2|i)$ to maintain conservation of probability. Typical values of N can range from 20 to 400, depending on the throughput rate.

Method Used to Determine the Equilibrium pmf π: The equilibrium pmf π (a row vector) must satisfy the following matrix equation:

$$\pi = \pi \mathbf{P}$$

where $\mathbf{P}$ is the transition probability matrix with elements $p_{ij} \triangleq Pr(R_{k+2} = j | R_k = i)$.

Instead of solving the N equations in N unknowns, it is computationally preferable to use the iterative procedure of relaxation, i.e.,

$$\pi = \lim_{n \to \infty} \pi(0)\mathbf{P}^n$$

where $\pi(0)$ is an arbitrary initial pmf. In all of our numerical examples we assumed that $\pi(0)$ is uniformly distributed between 0 and 9. A scaling operation, which ensures that the elements of the pmf sum to 1 at each iteration, must be included to minimize the effects of computer roundoff error.

The iteration was stopped when the following criterion was satisfied:

$$|\pi_j(n+1) - \pi_j(n)| \leq 0.5 \times 10^{-6}, \quad 0 \leq j \leq N-1.$$

We remark that the numerical considerations involved in this procedure represent a problem of interest for Markov chains in general, and have formed the basis of a paper that was presented at the First International Workshop on the Numerical Solution of Markov Chains [19].

Table 1 shows the size of the arrays used for PR-IFFO as a function of throughput, as well as the number of iterations needed for convergence for $M = 10$ (the number is somewhat smaller for smaller values of M and somewhat larger for larger values). To provide an indication of the behavior of the tail of the distribution, we also show the largest value of R_k that has a probability mass greater than 10^{-9}, which is denoted as $R_k(10^{-9})$. To verify that the array sizes are sufficiently large so that the effects of truncation are insignificant, additional computer runs have been made for larger array sizes. For the array sizes shown, all of the performance criteria considered (e.g., expected value and variance of R_k, expected delay, expected frame length, number of iterations required, and $R_k(10^{-9})$) are unchanged to at least the fourth decimal place when larger array sizes are used. In general, convergence and acceptable results are achieved for somewhat smaller array sizes.

Performance Results

Extensive performance results for the IFFO protocols, including comparisons with other multiple-access schemes, are presented in [4] and [5]. We now summarize some of these results. Fig. 4 shows delay-throughput curves for the three IFFO protocols that have been studied. At low to moderate throughput rates, F-IFFO provides considerable performance improvement over PR-IFFO; the ability to transmit in contention slots (thus avoiding the delay associated with the reservation process) has a significant impact on system operation. At high throughput rates, few packets are able to take advantage of the contention mode of operation, and PR-IFFO provides a close upper bound on expected delay under F-IFFO that becomes increasingly tight as throughput increases, as

TABLE 1
Size of Array Used, Number of Iterations Needed, and $R_k(10^{-9})$
as a Function of Throughput under PR-IFFO for $M = 10$

Throughput	Size of Array	Iterations	$R_k(10^{-9})$
0.00001	20	2	2
0.1	20	2	12
0.2	20	2	16
0.3	20	3	19
0.4	30	4	22
0.5	30	5	25
0.6	30	7	30
0.7	50	10	39
0.8	60	19	55
0.85	80	29	71
0.9	120	52	104
0.92	150	70	128
0.94	200	100	167
0.95	230	123	199
0.96	300	157	246
0.97	400	209	323

discussed earlier. Simulation results for PR-IFFO and F-IFFO were virtually identical to the numerically-computed results; this was expected because the mathematical model used to characterize these protocols is exact, except for the truncation of the pmf's. The CR*IFFO protocol (a special case of C-IFFO under which the transmission probability in contention slots depends on the system state and is chosen to optimize performance) provides only slightly better performance than F-IFFO. Simulation results are shown for CR*IFFO because they are more accurate than the numerically-computed results, since approximations were needed in the derivation of the transition probability matrix.

For all of the IFFO protocols, performance is quite insensitive to the number of

15

terminals for M greater than about 5; thus these curves for $M = 10$ are representative of higher values as well. Fig. 5 compares the delay-throughput performance of PR-IFFO and F-IFFO with that of TDMA for several values of M; perfect scheduling is also shown. The relative performance depends on M and throughput. Unlike the IFFO protocols, the performance of TDMA is highly sensitive to M. Fig. 6 compares the delay-throughput performance of F-IFFO with that of slotted ALOHA and several hybrid protocols that also combine random access with reservations. Of these protocols, F-IFFO performs best over a wide range of throughput (from 0.33 to 0.93 packets/slot).

These results demonstrate that hybrid protocols that combine features of reservation and contention can provide significantly improved performance over that of pure reservation schemes or pure contention-based schemes over a wide set of system parameters, while providing stable operation at high throughput rates. It is also significant to note that the performance of F-IFFO, which uses a very simple transmission policy in contention slots, performs almost as well as CR*IFFO, under which the transmission probabilities have been optimized.

7. THE INTERLEAVED-FRAME FIXED-LENGTH (IFFL) SCHEMES

The IFFO protocols are characterized by a frame length that adapts to channel traffic. This adaptive feature guarantees that all k-packets are successfully transmitted not later than the end of frame $k+2$; this is the flush-out feature of these schemes that, indeed, motivated them and that results in very high efficiency and excellent delay performance. In this section we consider a variation of these protocols under which the frame length is kept fixed at R slots. In certain applications (e.g., voice/data integration) it is desirable to keep a constant frame length, although doing so reduces the efficiency of the protocol. Clearly, for a fixed frame length of R slots, the maximum throughput that can be achieved is $(R-1)/R$ since one slot in each frame (the status slot) is needed

16

for overhead. In contrast, a throughput arbitrarily close to 1 can be achieved under the IFFO protocols.

We call these schemes the Interleaved-Frame Fixed-Length (IFFL) schemes. Operation is the same as that of IFFO, except that when R_k is greater than R, packets that cannot be accommodated in the current frame ($R_k - R + 1$ of them) are delayed until frame $k+2$, at which point they are again subject to further delays if there is again a large backlog. We refer to these packets as "excess packets." Clearly, these protocols violate the flush-out condition, and do not belong to the class of IFFO protocols. We may consider PR-IFFL and F-IFFL versions of these schemes, whose definitions follow from those given earlier for the IFFO schemes. System evolution is again characterized by the first-order Markov chain R_k. However, note that a slight reinterpretation of R_k is needed. It was previously defined as the number of reserved slots in frame k. Since the excess packets are delayed until frame $k+2$, it must now be interpreted as the number of packets for which reservations are needed at the beginning of frame k (including the excess packets, which have to be delayed). For PR-IFFL we have

$$R_{k+2} = A_k + \max(R_k - R + 1, 0).$$

By straightforward modification of the expressions for PR-IFFO, the elements of the transition probability matrix for PR-IFFL may now be written as

$$p_{ij} \triangleq Pr(R_{k+2} = j \mid R_k = i) = \begin{cases} \binom{MR}{j} \lambda^j (1-\lambda)^{MR-j}, & 0 \le i \le R-1 \\ \binom{MR}{j-i+R-1} \lambda^{j-i+R-1}(1-\lambda)^{MR-(j-i+R-1)}, & i \ge R-1. \end{cases}$$

For F-IFFL we observe that for $R_k < R-1$ the transition probabilities derived for F-IFFO apply, whereas for $R_k \ge R-1$ those for PR-IFFL must be used; hence the transition probabilities for F-IFFL are:

$$p_{ij} \triangleq Pr(R_{k+2} = j \mid R_k = i) = \begin{cases} \binom{M(i+1)}{j} \lambda^j (1-\lambda)^{M(i+1)-j} * c_{N_k}(j), & 0 \le i \le R-1 \\ \binom{MR}{j-i+R-1} \lambda^{j-i+R-1}(1-\lambda)^{MR-(j-i+R-1)}, & i \ge R-1. \end{cases}$$

Performance Results

Performance evaluation for the IFFL schemes proceeds in the same manner as that described earlier for the IFFO schemes. Fig. 7 shows the expected value of R_k under PR-IFFL as a function of input rate. Note that, since the frame length is fixed at R slots, the maximum throughput that can be achieved is $(R-1)/R$ because one slot in each frame (the status slot) is needed for overhead. We may define the "efficiency" achieved under the IFFL protocols to be the throughput multiplied by the ratio $R/(R-1)$. For example, for $R = 12$, a throughput rate of 0.9 packets/slot (the maximum shown in Fig. 7) corresponds to an efficiency of 0.9818. In contrast, under the IFFO protocols the input rate (and thus the throughput) can be arbitrarily close to 1 because the frame length expands to accommodate all reservations. Complete performance results for the IFFL schemes are not yet available.

8. NON-INTERLEAVED-FRAME FIXED-LENGTH (NIFFL) SCHEMES

Under the IFFO and IFFL schemes, the system state in even-numbered frames is independent of that in odd-numbered frames. Thus it is possible, e.g., for the even-numbered slots to build up large backlogs (high values of R_k) while the odd-numbered slots are lightly loaded. Since the IFFO schemes flush out all k-packets by the end of frame $k+2$, no inefficiency arises from this behavior. However, under IFFL, whenever $R_k > R-1$ the excess packets (R_k-R+1 of them, as discussed earlier) will be postponed to frame $k+2$. If there are some unreserved slots in frame $k+1$, it would be advantageous to transmit some or all of these excess packets in those slots.

To address this situation we consider a variation of the IFFL protocols that we call the Non-Interleaved-Frame Fixed-Length (NIFFL) protocols. We show that the PR-NIFFL version is again characterized by an underlying first-order Markov chain, and thus can be evaluated using techniques similar to those used for the IFFO schemes. However, the description of F-NIFFL requires a second-order Markov chain, which

makes performance evaluation considerably more difficult. The system evolution for the NIFFL protocols may be described as follows:

$$R_{k+2} = R_{k+2}^{(k)} + R_{k+2}^{(k+1)}.$$

$R_{k+2}^{(k)}$ is the number of k-packets that are included in R_{k+2}, i.e., all arrivals in frame k, except those that were transmitted successfully in the contention slots of frame k. Thus we have

$$R_{k+2}^{(k)} = A_k - S_k.$$

$R_{k+2}^{(k+1)}$ is the number of excess packets that are carried over from frame $k+1$. We have

$$R_{k+2}^{(k+1)} = \max\left\{[R_{k+1} - (R-1)], 0\right\}.$$

The Markov chain now is the pair (R_{k+1}, R_k), and thus we need

$$p_{ij \to jm} \triangleq Pr(R_{k+2} = m, R_{k+1} = j \mid R_{k+1} = j, R_k = i).$$

A brute-force system description would require a state probability vector $Pr(R_{k+1}, R_k)$ of dimension N^2, where N is the truncation value as discussed earlier. We can write

$$Pr(R_{k+1}, R_k) \triangleq \begin{pmatrix} \mathbf{q}_0 & \mathbf{q}_1 & \cdots & \mathbf{q}_{N-1} \end{pmatrix}$$

where $\mathbf{q}_j$ is the row vector whose entries are

$$q_{ji} \triangleq Pr(R_{k+1} = j, R_k = i).$$

Thus a transition probability matrix of size $N^2 \times N^2$ would be needed. The dimensions of the problem can be reduced somewhat by making the observation that R_{k+1} includes all of the excess packets contained in R_k. Thus, given R_{k+1}, the exact value of R_k is needed only if it is less than $R-1$. We define an aggregate state $R_k = R-1$ that actually contains all states for which $R_k \geq R-1$. Now the state (R_{k+1}, R_k) can be described by a vector of dimension NR rather than N^2. The state (R_{k+2}, R_{k+1}) still requires a vector of dimension N^2, however. The resulting transition probability matrix is of dimension

$NR \times N^2$. This is still of unmanageable size. Before discussing a method to decompose the problem into one that requires N matrices of dimension $R \times N$ (still a very large size, but considerably smaller than before), we demonstrate how the PR-NIFFL protocol can be evaluated using a first-order Markov chain.

Under PR-NIFFL, $R_{k+2}^{(k)} = A_k$ because $S_k = 0$ (since there are no contention transmissions in a pure reservation system). Thus

$$R_{k+2} = A_k + \max\left\{[R_{k+1} - (R-1)], 0\right\}.$$

Since the frame length is constant, A_k does not depend on R_k. It is binomially distributed with parameter λ over MR trials. Thus the system description for PR-NIFFL can be reduced to a first-order Markov chain as follows:

$$p_{ij} \triangleq Pr(R_{k+2} = j \mid R_{k+1} = i) = \begin{cases} \binom{MR}{j} \lambda^j (1-\lambda)^{MR-j}, & 0 \le i \le R-1 \\ \binom{MR}{j-i+R-1} \lambda^{j-i+R-1} (1-\lambda)^{MR-(j-i+R-1)}, & i \ge R-1. \end{cases}$$

Note that this expression is identical to that for PR-IFFL, except that R_k is replaced here by R_{k+1}.

The system evolution of F-NIFFL cannot be described by a first-order Markov chain. However, the second-order Markov chain can be decomposed to generate a collection of smaller problems, as we now discuss. We start from the observation made earlier that R_{k+1} is common to both states (i.e., origin and destination) of each transition. This led to the state probability description in terms of the vectors of the form $\mathbf{q}_j$ with elements q_{ji}; the index j takes on all possible values of R_{k+1} (i.e., from 0 to $N-1$) and i takes on all possible values of R_k (i.e., from 0 to $R-1$).* For each value of R_{k+1} we consider the transitions from R_k to R_{k+2}. The corresponding transition probabilities are denoted as

---

* As noted earlier, the state $R_k = R-1$ is an aggregate state that actually contains all values of $R_k \ge R-1$.

$$p_i \xrightarrow{j} m \triangleq p_{ij \to jm} \triangleq Pr(R_{k+2} = m, R_{k+1} = j \mid R_{k+1} = j, R_k = i).$$

Whenever $j \leq R-1$, $R_{k+2}^{(k+1)} = 0$, in which case $R_{k+2} = R_{k+2}^{(k)}$. Whenever $j > R-1$, $R_{k+2}^{(k+1)} = j - (R-1)$ packets must be added to $R_{k+2}^{(k)}$. By an appropriate modification of the expressions previously derived for F-IFFL, we have the following transition probabilities for F-NIFFL:

$$p_i \xrightarrow{j} m = \begin{cases} \binom{M(i+1)}{m} \lambda^m (1-\lambda)^{M(i+1)-m} * c_{N_k}(m), & 0 \leq j \leq R-1 \\ \binom{M(i+1)}{m-j+R-1} \lambda^{m-j+R-1}(1-\lambda)^{M(i+1)-(m-j+R-1)} * c_{N_k}(m), & j > R-1. \end{cases}$$

Recall that we must have $i \leq R-1$ since $R-1$ is an aggregate state (the excess packets are incorporated into $j = R_{k+1}$). Again, the convolution vanishes whenever $i = R-1$. Note that the two expressions given here correspond to different ranges of j, rather than i. There are N transition probability matrices of this type (one for each value of j, denoted $\mathbf{P}_{(j)}$), each of dimension $R \times N$. Each $\mathbf{q}_j$ vector is multiplied by the corresponding transition probability matrix $\mathbf{P}_{(j)}$. The result is again a collection of N probability vectors for R_{k+2}, each of dimension N, one for each value of R_{k+1}. The vector corresponding to $R_{k+1} = j$, which we denote $\mathbf{r}_j$, consists of the elements

$$r_{jm} = Pr(R_{k+1} = j, R_{k+2} = m).*$$

In preparation for the next iteration, the state probabilities are rearranged in the form of the $\mathbf{q}_j$ vectors, where all states for which $R_{k+1} \geq R-1$ are combined in the aggregate state $R-1$. This procedure is repeated until convergence is achieved.

In our discussion of the IFFO protocols we made the observation that the performance of PR-IFFO bounds closely that of F-IFFO in the limit of high throughput rates. Similarly, at high throughputs the performance of PR-NIFFL provides a tight upper bound on that of F-NIFFL. This is true because most frames have no unreserved

---

* Note that $\mathbf{q}_j$ refers to a probability vector in which the state in the later of two slots is held constant (i.e., $R_{k+1} = j$, while R_k varies), whereas $\mathbf{r}_j$ refers to a probability vector in which the state in the earlier of two slots is held constant (i.e., $R_{k+1} = j$, while R_{k+2} varies).

slots at high throughput rates, in which case $R_k = R-1$ (the aggregate state). Thus no k-packets can be transmitted in frame k, in which case F-NIFFL functions in the same manner as PR-NIFFL. Therefore, at high throughput rates (for which large transition probability matrices are needed), the performance of F-NIFFL can be bounded and closely approximated by that of PR-NIFFL, which is characterized by a much simpler description (i.e., a first-order Markov chain).

9. AN IFFL PROTOCOL FOR INTEGRATED VOICE/DATA SYSTEMS

The communication systems that gave rise to the models discussed thus far in this report can be modified slightly to handle voice traffic in addition to data packets. This is an important extension in communication system design, and represents our main objective here. A customary model for voice assumes that voice calls are generated at idle terminals according to a Bernoulli process, and that they are geometrically distributed in length; thus the probability that a call is completed in any particular frame is also a Bernoulli process. There are M_V voice users in the system and M data users. The time constants associated with voice traffic are considerably larger than those associated with data traffic; voice calls will typically last from tens to hundreds of frames. This difference in time durations plays a key role in the development of an approximate system model for this protocol.

To accommodate the needs of both voice and data traffic, we consider a channel-access protocol under which a reservation scheme is used for voice traffic and IFFL (which combines reservation and contention) is used for data.* We call these the Voice/Data IFFL (VD-IFFL) protocols. Under these schemes, once a voice call is accepted by the system, it is guaranteed access to one slot each frame until its completion.** The standard idea of a "movable-boundary" mechanism is used to

---

* Although better performance can be expected under the NIFFL schemes, we consider IFFL schemes here because they are easier to model.
** Here we implicitly assume that the slot length has been selected in conjunction with the frame length (which is equal to the propagation delay) so that the voice burst rate results in

partition each frame between voice and data operation, as shown in Fig. 8. A fixed-length frame structure is necessary to accommodate the real-time requirements of voice traffic.***

Voice calls are accepted by the system as long as the number of calls does not exceed some maximum value V_{max}, which must be less than R. If a slot is not available for a new call, the call is assumed to be lost; there is no buffering of voice calls. The slots not used by voice calls (including empty slots in the voice portion of the frame) are used for data traffic, which is transmitted by using one of the IFFL protocols. Since the decision to accept new voice calls depends only on whether or not the threshold V_{max} is exceeded, voice traffic is unaffected by data traffic; however, the operation of the data protocol is dependent on voice traffic because data traffic is permitted to use unneeded slots in the voice portion. Thus this problem is similar to variable service rate queueing systems in which the service rate depends on another process. Our model can be extended to consider systems in which the decision to accept new voice calls also depends on the system backlog (i.e., the value of R_k). However, the analysis of such systems is considerably more difficult and is not addressed here. Multiplexing (although not channel access) systems incorporating this feature were studied by Viniotis and Ephremides [20, 21].

As shown in Fig. 8, the first slot of every frame is once again the status slot, during which each terminal transmits its reservations for packets that arrived in the previous slot. The next V_k slots are reserved for voice traffic, where V_k is the number of voice calls in progress at the beginning of slot k. The remainder of the frame consists of D_k

---

the required symbol rate needed for real-time voice transmission. We could actually use any frame length greater than or equal to the propagation delay such that the voice transmission rate is correct.
*** A fixed frame length is not necessary if real-time delivery of voice traffic is not required. In fact, in certain military applications that are characterized by poor speech quality, delays of several hundred ms or more may be considered to be acceptable if sufficient buffer space is available at the destination to permit the reconstruction of a continuous voice stream. In such cases, the IFFO protocols (or perhaps a variant of them with a constraint on maximum frame length) may provide acceptable performance. The study of the delay characteristics of voice in systems with variable frame lengths is an interesting topic for future research.

data slots, where

$$D_k = R - 1 - V_k.$$

As with the protocols designed purely for data, R_k is the number of data packets for which reservations are needed at the beginning of frame k. Whenever $R_k < D_k$, N_k^v slots are available for contention transmission, where

$$N_k^v = \max(D_k - R_k, 0) = \max(R - 1 - v - R_k, 0)$$

for each particular value of $V_k = v$. Whenever $R_k > D_k$, the excess packets are delayed until frame $k+2$. Operation of the data portion of VD-IFFL can thus be viewed as that of IFFL with a variable number of slots (D_k) available for data traffic, where D_k depends on V_k. In contrast, under IFFL exactly $R-1$ slots are available for data in each frame.

An Exact Markov Chain Model for VD-IFFL

The development of a Markov chain model for the VD-IFFL protocols has taken into account two particular complicating features of the protocols, i.e., the dependence of data traffic on voice (whereas voice is independent of data) and the fact that the voice process requires a second-order Markov chain description (as is discussed shortly). Thus, the protocol can be characterized by the Markov chain (R_k, V_k, V_{k-1}), which has transition probabilities $Pr(R_{k+2}, V_{k+2}, V_{k+1}, | R_k, V_k, V_{k-1})$. A brute-force approach would consider a probability vector containing all possible triplets of V_k, V_{k-1}, and R_k. The maximum value of V_k and V_{k-1} would be the threshold value V_{max}; R_k would have a maximum value of $N-1$ as in the evaluation of the data-only IFFL protocols. Thus a transition probability matrix of dimension $(V_{max}+1)^2 N \times (V_{max}+1)^2 N$ would be needed (e.g., for $V_{max} = 6$ and $N = 100$, these matrices would be 4900 × 4900.) However, not all transitions are possible, and dramatic reductions in the number of computations needed can be made by decomposing the problem into separate voice and data portions. We do

this by recognizing that the voice-call process does not depend on the data-message process.* Thus

$$Pr(R_{k+2},V_{k+2},V_{k+1}|R_k,V_k,V_{k-1}) = Pr(R_{k+2}|R_k,V_k) \, Pr(V_{k+2},V_{k+1}|V_k,V_{k-1}).$$

The transition from R_k to R_{k+2} depends on V_k (because V_k determines D_k), but not on V_{k+1} or V_{k+2}. The transition from (V_k,V_{k-1}) to (V_{k+2},V_{k+1}) does not depend on R_k or R_{k+2}. We first note that these observations simplify the evaluation of the transition probability matrix. Actually, a much greater benefit is realized. It is demonstrated below that the transitions corresponding to the data process can be considered separately for each value of V_k. Thus, it is not necessary to perform the iteration with the huge transition probability matrix that characterizes the evolution of the complete voice/data state description. The evaluation of system performance can be decomposed into a number of smaller problems that are of manageable size.

Before proceeding, we simplify the notation by defining $\overline{V}_k = (V_k,V_{k-1})$. This yields

$$Pr(R_{k+2},\overline{V}_{k+2}|R_k,\overline{V}_k) = Pr(R_{k+2}|R_k,V_k) \, Pr(\overline{V}_{k+2}|\overline{V}_k).$$

We recognize that the transitions from frame k to frame $k+2$ can be modeled as a two-step process. Data transitions are considered first. Given $Pr(R_k,\overline{V}_k)$, we first determine $Pr(R_{k+2},\overline{V}_k)$. This requires a different transition probability matrix for each value of V_k, as explained below (note that V_{k-1} does not affect these transitions). This operation can be expressed as follows:

$$Pr(R_{k+2} = j, \overline{V}_k = \overline{v}) = \sum_{i=0}^{N} Pr(R_{k+2}=j|R_k=i,\overline{V}_k=\overline{v}) \, Pr(R_k=i,\overline{V}_k=\overline{v}) \quad 0 \leq j \leq N-1.$$

Next, the voice transitions are considered. Given $Pr(R_{k+2},\overline{V}_k)$, we determine $Pr(R_{k+2},\overline{V}_{k+2})$. Since the voice transitions are independent of the data traffic, the same transition probability matrix is used for all values of R_{k+2}. Thus, the following is

---

* This is not true for systems in which the decision on whether or not to accept a voice call is permitted to depend on R_k, however.

25

evaluated:

$$Pr(R_{k+2} = j, \overline{V}_{k+2} = \overline{w}) = \sum_{\overline{v}=(0,0)}^{(V_{max}, V_{max})} Pr(\overline{V}_{k+2} = \overline{w} \,|\, \overline{V}_k = \overline{v}) \, Pr(R_{k+2} = j, \overline{V}_k = \overline{v})$$

The equilibrium distribution of the system state is determined by repeating this two-step iteration until convergence is achieved. We emphasize that this model is exact; the reduction in the size of the transition matrices has been achieved by exploiting specific structural properties of the Markov chain.

Data Transitions:

We first consider the data transitions. Corresponding to each value of V_k there is an $N \times N$ transition probability matrix for the data message process with elements

$$p_{ij}^v \triangleq Pr(R_{k+2} = j \,|\, R_k = i, V_k = v).$$

These transition probabilities are easily obtained from those for the IFFL protocols. Under IFFL, $R-1$ slots are available for packet transmission in each frame. Under VD-IFFL, this number is reduced to $D_k = R-1-V_k$. Thus for each value of $V_k = v$ we replace $R-1$ by $R-1-v$. This yields for PR-VD-IFFL:

$$p_{ij}^v = \begin{cases} \binom{MR}{j} \lambda^j (1-\lambda)^{MR-j}, & 0 \le i \le R-1-v \\ \binom{MR}{j-i+R-1-v} \lambda^{j-i+R-1-v} (1-\lambda)^{MR-(j-i+R-1-v)}, & i \ge R-1-v. \end{cases}$$

Similarly, for F-VD-IFFL we obtain

$$p_{ij}^v = \begin{cases} \binom{M(i+1)}{j} \lambda^j (1-\lambda)^{M(i+1)-j} * c_{N_k^v}(j), & 0 \le i \le R-1-v \\ \binom{MR}{j-i+R-1-v} \lambda^{j-i+R-1-v} (1-\lambda)^{MR-(j-i+R-1-v)}, & i \ge R-1-v. \end{cases}$$

Thus, at each iteration, $(V_{max} + 1)$ matrix multiplications (each of size $N \times N$, one for each value of v) must be carried out to determine the data transitions.

Voice Transitions:

Next, we consider the voice transitions. The probability that a new call is generated at an idle terminal during any particular frame is denoted as λ_V. The probability that an ongoing call completes service during any particular frame is denoted as μ_V. Reservations are made during the status slot, in the same manner as those for data traffic. Thus, if a voice call arrives during frame k, its reservation will be transmitted during the first slot of frame $k+1$, and, if available, a slot will be reserved for it beginning in frame $k+2$. A voice call will be accepted only if doing so would not raise the number of voice calls in the system to a number greater than the threshold V_{max}. When a call is blocked (because the threshold has been reached), it is dropped from the system and the terminal reenters the idle state. An end-of-message (EOM) indicator is transmitted at the end of the last packet in the call to indicate that the slot is no longer needed in subsequent frames. Since the frame length is assumed to be equal to the propagation delay, an EOM transmitted in frame k is not received until the corresponding slot of frame $k+1$; thus this slot remains idle during frame $k+1$ and becomes available to a new voice call in frame $k+2$.*

It is important to note that there is no frame interleaving of the voice call process (other than that associated with making the reservation and receiving the EOM), since a call occupies a time slot in every frame from its start to its completion. Thus the transition from frame k (which is characterized by $\overline{V}_k$) to frame $k+2$ (which is characterized by $\overline{V}_{k+2}$) represents a two-phase transition. The evolution of the voice process in one such phase proceeds as follows:

$$V_{k+2} = \min\left\{ V_{k+1} + A_k^V - U_k, V_{max} \right\}$$

---

* It might be possible to avoid the wasted slot in frame $k+1$ by transmitting the EOM in the status slot of frame k, since knowledge that the call is terminating may be available at that time. Incorporating such a feature into the model would require a slight modification to the description of the voice transition process. Alternatively, if the frame length were constrained to be at least $R+1$ slots, then knowledge of frame-k departures would be available in time to use the slot in frame $k+1$.

where A_k^V is the number of voice arrivals in frame k and U_k is the number of voice calls that are completed in frame k. Note that a second-order Markov chain is needed to describe the voice process; i.e., we need the transition probabilities of the form $Pr(\overline{V}_{k+2}|\overline{V}_k) = Pr(V_{k+2},V_{k+1}|V_{k+1},V_k)$. We approach this problem by defining the transition probability matrix $\mathbf{S}$ of size $(V_{\max} + 1)^2 \times (V_{\max} + 1)^2$. Note that no truncation is required because the number of voice calls cannot exceed $V_{\max}$. Since the matrix is of easily manageable size (e.g., 49×49 for $V_{\max} = 6$), it is not necessary to use a decomposition of the form presented earlier for F-NIFFL.

To obtain the transition probability from $\overline{V}_k$ to $\overline{V}_{k+2}$, we define the two-step transition probability matrix $\mathbf{S}^{(2)}$, where, clearly, $\mathbf{S}^{(2)} = \mathbf{S}^2$.

To evaluate the elements of $\mathbf{S}$, we first consider the arrival process in frame k. Each of the currently idle users ($M_V - V_k$ of them) will generate a new call with probability λ_V. Thus

$$Pr(A_k^V = j \,|\, V_k = i) \;=\; \binom{M_V - i}{j} \lambda_V^j \,(1 - \lambda_V)^{M_V - i - j}.$$

Now consider the V_k active voice users (not including the new additions), at each of which a call will be completed with probability μ_V. We have

$$Pr(U_k = j \,|\, V_k = i) \;=\; \binom{i}{j} \mu_V^j \,(1 - \mu_V)^{i-j}.$$

A_k and U_k are conditionally independent, given V_k. Thus it is straightforward, although somewhat tedious, to obtain the elements of the transition probability matrix $\mathbf{S}$ (and therefore $\mathbf{S}^{(2)}$) for the voice transitions.

Approximate Models for the VD-IFFL Protocols

Despite the simplifications to the model that have been possible as a result of the structural properties of the VD-IFFL protocols, their performance evaluation remains computationally intensive. Thus it is desirable to develop approximate models that will

provide a good estimate of system performance. We consider two such approximations, which can be used simultaneously. The first is based on the markedly different time scales of voice and data traffic, i.e., on the fact that the voice process changes very slowly in comparison to the data process. In the second, the second-order Markov chain model is replaced by an approximate first-order Markov chain.

A Quasi-Static Model

Since the interarrival times and holding times of voice calls are considerably longer than a slot duration, V_k is a slowly changing quantity. The system may be described by a number of "super-states," each of which is characterized by a different value of V_k, and which contain all possible values of R_k. First, we assume that V_k is constant for all time. For each possible value of V_k we determine the equilibrium distribution of R_k by using the transition probabilities p_{ij}^v discussed earlier. Then we average the distribution of R_k over all possible values of V_k, whose distribution can be determined exactly following the procedure we just discussed. Courtois' concept of near-complete decomposability [22] appears to be applicable here.

This approach appears to be reasonable because the voice-call transitions are totally decoupled from data-message transitions, although the converse is not true in the sense that the data-message transition probabilities depend on V_k. It is expected that agreement between the approximate and exact models will be close in the limit of very long interarrival and holding times. However, computational results are not available yet to confirm this expectation.

It is difficult to predict with certainty whether this approximate model will result in significantly reduced computation time. If the average number (over all values of V_k) of iterations needed using the approximate model is less than the number of iterations under the exact model, improvement will be realized. We anticipate that the average number of iterations will be reduced when V_k is held fixed because transitions for different values of V_k are completely decoupled from each other. Again, this plausible

29

argument needs to be confirmed via computational results in the future. Another benefit of the approximate model is a reduced storage requirement because only one $N \times N$ matrix is used at any time (instead of $V_{max} + 1$ of them).

A First-Order Markov Chain Model for the Voice Process

As noted earlier, a second-order Markov chain is needed to describe the voice process under the VD-IFFL protocols because V_{k+2} depends on V_{k+1}, A_k^V and U_k. We already made the observation that a slight modification to the protocol (either incorporation of an EOM in the status slot of the frame in which the call terminates, or alternatively the use of frame lengths of at least $R+1$ slots) would permit V_{k+2} to depend on U_{k+1} rather than on U_k. However, V_{k+2} would still depend on A_k^V. One way around this difficulty is to make another quasi-static assumption. In this case, we assume that the distribution of the number of arrivals in frame $k+1$ is the same as that in frame k. Again, this should be reasonable because the voice process changes slowly, as compared to the data process. Such an approximation is expected to be fairly accurate, especially in the limit of a large number of users (in which case the arrival process is less dependent on the number of ongoing calls) and in the limit of long voice calls. The advantage of the use of a first-order Markov chain model is that it would permit a state description of the voice process in terms of a vector of dimension $(V_{max}+1)$ instead of $(V_{max}+1)^2$. The accuracy of this approximation will be evaluated in the near future.

10. OPERATION IN A GROUND-RADIO ENVIRONMENT

Satellite channels are characterized by a round-trip propagation delay of approximately 0.27 second, which may typically correspond to $R = 10$ to 12 slots. In contrast, in ground radio channels the propagation delays are typically measured in ms, which correspond to $R \ll 1$. In this case, the interleaved frame structure is not needed because reservations are always received prior to the start of the next frame. Data-only protocols can be described by a first-order Markov chain that represents transitions from

R_k to R_{k+1}.* In particular, the complex second-order Markov chain descriptions needed to characterize the NIFFL protocols are no longer needed because the transition probabilities for the IFFL protocols now correspond to transitions from frame k to frame $k+1$ under NIFFL. For example, in [17] a Pure-Reservation Direct-Flush-Out (PR-DFO) scheme was discussed, and simple expressions were derived for expected system delay.

When $R \ll 1$, there is normally no reason to incorporate contention operation into the protocol because reservations can be made in the next unreserved slot. However, when considering integrated voice/data systems, it is appropriate to maintain a fixed frame length because each packet of a voice call needs a periodically recurring time slot. Thus it may be appropriate to maintain frame lengths of at least some specified length even when the propagation delay is near zero. The transition probabilities for VD-NIFFL protocols for the case of very small propagation delay are obtained directly from those for VD-IFFL. The only difference is that the voice transitions are now characterized by a first-order Markov chain since (like data) reservations are received prior to the beginning of the next frame and because knowledge of call completions is available prior to the next frame. Also, voice transitions now represent the single step from frame k to frame $k+1$.

11. A FRAMED-ALOHA PROTOCOL FOR INTEGRATED VOICE/DATA SYSTEMS

The movable-boundary approach can be used to adapt a variety of framed channel access protocols, which were originally developed for data-only operation, to integrated networks. For example, we may consider the Framed-ALOHA protocol [6]. In the simplest version of this protocol, time is divided into fixed length frames. In each frame, every non-empty terminal transmits a packet in a slot chosen at random. Unsuccessful packets are retransmitted in the next frame. In satellite networks an

---

* In integrated voice/data protocols, data transmissions also depend on V_k.

interleaved-frame version of the protocol may be used for the reasons discussed earlier for the IFFO protocols. A first-order Markov chain is used to describe system dynamics.

Under integrated operation, V_k slots are needed for voice traffic in frame k. As before, voice traffic uses a reservation mechanism for channel access. The remaining slots in the fixed-length frame are available for data traffic, as discussed earlier for VD-IFFL. Thus the data portion of the protocol operates as a Framed-ALOHA scheme with a variable frame length, which is equal to the number of slots available for data (which of course depends on the number used for voice); however, data packets may arrive in any slot, including those assigned to voice. It is straightforward to incorporate a variable frame length into the analysis of this protocol. The analysis for voice is identical to that presented in Section 9; when no frame interleaving is used, a one-step transition is sufficient.

For the cases of no capture and perfect capture, expressions can be derived for the transition probabilities for system backlog. However, for the more interesting case of a general capture model, for which a powerful combinatorial technique was developed in [6], the transition probabilities must be evaluated numerically. In any of these cases, evaluation of the equilibrium performance requires a numerical solution similar to that done for the IFFO and related protocols.

The performance of Framed ALOHA in an integrated environment will be studied in the future if time permits. We have included this brief discussion here to show that the movable-boundary method can be applied to a wide variety of data protocols to extend their applicability to integrated networks.

12. CONCLUSIONS

In this report we have addressed the major issues associated with channel access in integrated radio networks, and we have derived the transition probabilities associated

with one particular class of protocols that we have developed for this application, namely the Voice/Data Interleaved-Frame Fixed-Length (VD-IFFL) protocols. The VD-IFFL protocols are a modification of the Interleaved-Frame Flush-Out (IFFO) protocols, which were originally developed for data-only applications. As a first step toward the study of protocols for integrated networks, we considered the IFFL protocols (also for data only), which are similar to the IFFO protocols, except that under IFFL the frame length is kept fixed, a characteristic that is desirable in applications involving voice traffic. It was then straightforward to incorporate voice traffic into the protocol by using a movable-boundary scheme and to obtain the transition probabilities for this new scheme.

This report has emphasized geosynchronous satellite networks because their analysis is more challenging than that of ground-radio networks. Geosynchronous satellite networks are characterized by a round-trip propagation delay of about 0.27 second. The interleaved frame structure was introduced to accommodate the impact of this long delay on the reservation process. In contrast, ground-radio networks are characterized by delays of milliseconds, and frame interleaving is not needed. The transition probabilities for the non-interleaved schemes are easily obtained from those for the interleaved schemes.

The IFFO and IFFL protocols are characterized by an infinite first-order Markov chain. Clearly, truncation is needed to evaluate equilibrium system performance. Otherwise, the system models that have been developed are exact. Performance can be evaluated by the iterative procedure of relaxation. VD-IFFL is described by a two-dimensional Markov chain that is first order in terms of the data state and second order in terms of the voice state. We showed how the problem could be broken up into a number of smaller problems; at each iteration, transition probabilities have been derived for the data-message process for each possible number of voice calls in the system. Each data-message transition of this type is then followed by a voice-call transition. We

33

also presented an approximate solution method that takes advantage of the different time scales associated with the voice and data processes; i.e., a voice call will generally last tens to hundreds of slots, whereas data consists of individual packets.

We also considered the Non-Interleaved-Frame Fixed-Length (NIFFL) schemes, which will provide better performance than the IFFL schemes because packets for which reservations have been received do not have to be delayed an extra frame. These schemes (when used in satellite networks) are characterized by a two-dimensional Markov chain, which complicates the numerical evaluation process greatly because of the large number of states that can be generated. We showed that the pure-reservation version of this protocol (PR-NIFFL) can be described by a first-order Markov chain. For the case of the Fixed-Contention version (F-NIFFL), for which the two-dimensional representation is needed, we demonstrated how this problem can be broken up into a number of smaller problems each of which is similar to that associated with the IFFO protocols.

At this time, numerical results are available only for the IFFO protocols. In the near future we expect to evaluate the performance of the IFFL and VD-IFFL schemes as well. We also plan to develop optimization models for integrated voice/data protocols. Thus far, we have considered only schemes in which the decision to accept a voice call is based simply on whether or not the threshold V_{max} has been reached. In the future we plan to consider systems in which the decision to accept or block a call is based on a weighted sum of voice-call blocking probability and data-message delay. Markovian decision process models appear to be a reasonable approach for this optimization problem. Thus far, these techniques have been applied to the movable-boundary schemes for multiplexing at a single node and for simple tandem configurations [20, 21], but not to the channel access problem.

References

1. F. A. Tobagi, "Multiaccess Protocols in Packet Communication Systems," *IEEE Transactions on Communications* **COM-28** pp. 468-488 (April 1980).

2. S. S. Lam, "Multiple Access Protocols," in *Computer Communications, Volume I: Principles,* ed. W. Chou, Prentice Hall (1983).

3. J. E. Wieselthier and A. Ephremides, "Voice/Data Integration in Mobile Radio Networks: Overview and Future Research Directions," NRL Report 9189, Naval Research Laboratory, Washington, D.C. (September 1989).

4. J. E. Wieselthier and A. Ephremides, "A New Class of Protocols for Multiple Access in Satellite Networks," *IEEE Transactions on Automatic Control* **AC-25** pp. 865-879 (October 1980).

5. J. E. Wieselthier, "A New Class of Multi-Access Protocols for Packet Communication Over a Satellite Channel—The Interleaved Frame Flush-Out Protocols," Ph.D. Dissertation, University of Maryland (March 1979).

6. J. E. Wieselthier, A. Ephremides, and L. A. Michaels, "An Exact Analysis and Performance Evaluation of Framed ALOHA With Capture," *IEEE Transactions on Communications* **37** pp. 125-137 (February 1989).

7. L. G. Roberts, "Dynamic Allocation of Satellite Capacity Through Packet Reservation," *AFIPS Conference Proceedings, 1973 National Computer Conference* **42** pp. 711-716 (1973).

8. K. Kummerle, "Multiplexer Performance for Integrated Line- and Packet-Switched Traffic," *Proceedings of the Second International Conference on Computer Communication (ICCC)* pp. 508-515 (1974).

9. P. Zafiropoulo, "Flexible Multiplexing for Networks Supporting Line-Switched and Packet-Switched Data Traffic," *Proceedings of the Second International Conference on Computer Communication (ICCC)* pp. 517-523 (1974).

10. G. J. Coviello and P. A. Vena, "Integration of Circuit/Packet Switching by a SENET (Slotted Envelope Network) Concept," *Conference Record of National Telecommunications Conference* pp. 42.12-42.17 (1975).

11. T. Suda, H. Miyahara, and T. Hasegawa, "Performance Evaluation of an Integrated Access Scheme in a Satellite Communication Channel," *IEEE Journal on Selected Areas in Communications* **SAC-1** pp. 153-164 (January 1983).

12. C.-S. Wu and V. O. K. Li, "Integrated Voice/Data Protocols for Satellite Channels," *Conference Record of Mobile Satellite Conference, Jet Propulsion Laboratory* pp. 413-422 (May 1988).

13. J. E. Wieselthier, "Code Division Multiple Access Techniques and Their Application to the High-Frequency (HF) Intratask Force (ITF) Communication Network," NRL Report 9094, Naval Research Laboratory, Washington, D.C. (September 1988).

14. M. Soroushnejad and E. Geraniotis, "Voice/Data Integration in Random-Access Code Division Multiple-Access Packet Radio Networks," *Proceedings of the Twenty-Third Annual Conference on Information Sciences and Systems (CISS), Johns Hopkins University* pp. 706-710 (March 1989).

15. M. Soroushnejad and E. Geraniotis, "Performance Evaluation of Multi-Access Strategies for an Integrated Voice/Data CDMA Packet Radio Network," *presented at the 1990 IEEE International Symposium on Information Theory* (January 1990).

16. E. Geraniotis and M. Soroushnejad, "Performance Evaluation of Multi-Access Strategies for an Integrated Voice/Data Spread-Spectrum Packet Radio Network," submitted as an NRL Memorandum Report, Naval Research Laboratory, Washington, D.C. (February 1990).

17. J. E. Wieselthier and A. Ephremides, "Protocols of Multiple Access (A Survey) and the IFFO Protocols (A Case Study) – Invited," *Proceedings of the NATO Advanced Study Institute on "New Concepts in Multi-User Communication, Norwich, U. K. 1980,* Sijthoff and Noordhoff, Alphen aan den Rijn, The Netherlands (1981).

18. F. G. Foster, "On the Stochastic Matrices Associated with Certain Queueing Processes," *Ann. Math. Statist.* **24** pp. 355-360 (1953).

19. J. E. Wieselthier and A. Ephremides, "Some Markov Chain Problems in the Evaluation of Multiple-Access Protocols," *Proceedings of the First International Workshop on the Numerical Solution of Markov Chains* pp. 258-282 (January 1990).

20. I. Viniotis and A. Ephremides, "Optimal Switching of Voice and Data at a Network Node," *Proceedings of the 28th Conference on Decision and Control (CDC)* pp. 1504-1507 (December 1987).

21. I. Viniotis and A. Ephremides, "On the Optimal Dynamic Switching of Voice and Data in Communication Networks," *Proceedings of the Computer Networking Symposium* pp. 8-16 (April 1988).

22. P. J. Courtois, *Decomposability: Queueing and Computer System Applications*, Academic Press (1977).
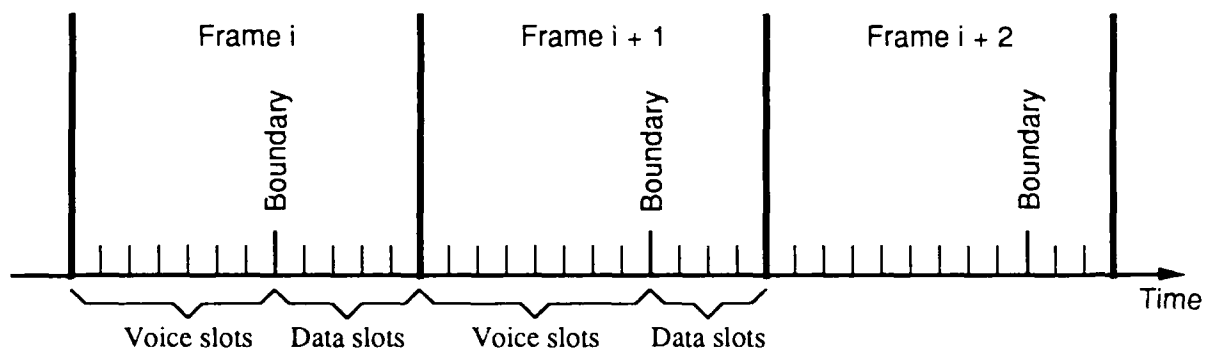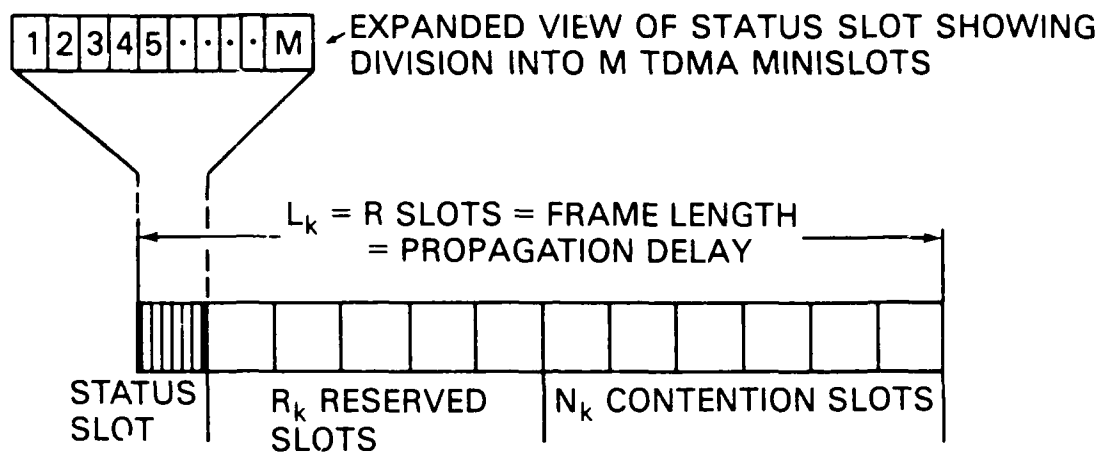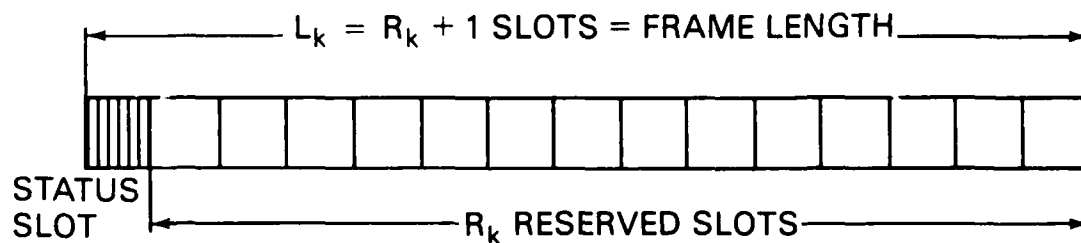
Fig. 1 The Movable-Boundary channel

CASE 1:   THE CASE WITH CONTENTION SLOTS, i.e., $R_k \leq R-2$

CASE 2:   THE CASE WITH NO CONTENTION SLOTS, i.e., $R_k \geq R-1$
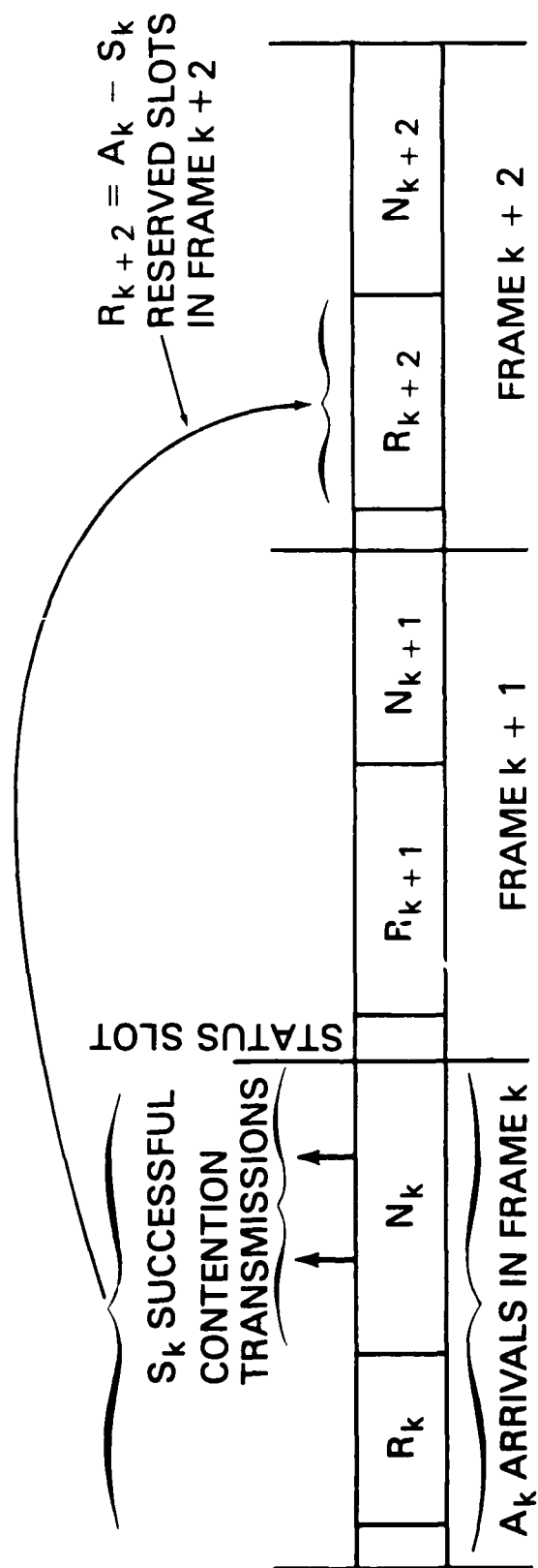
Fig. 2   Frame structure for the IFFO protocols

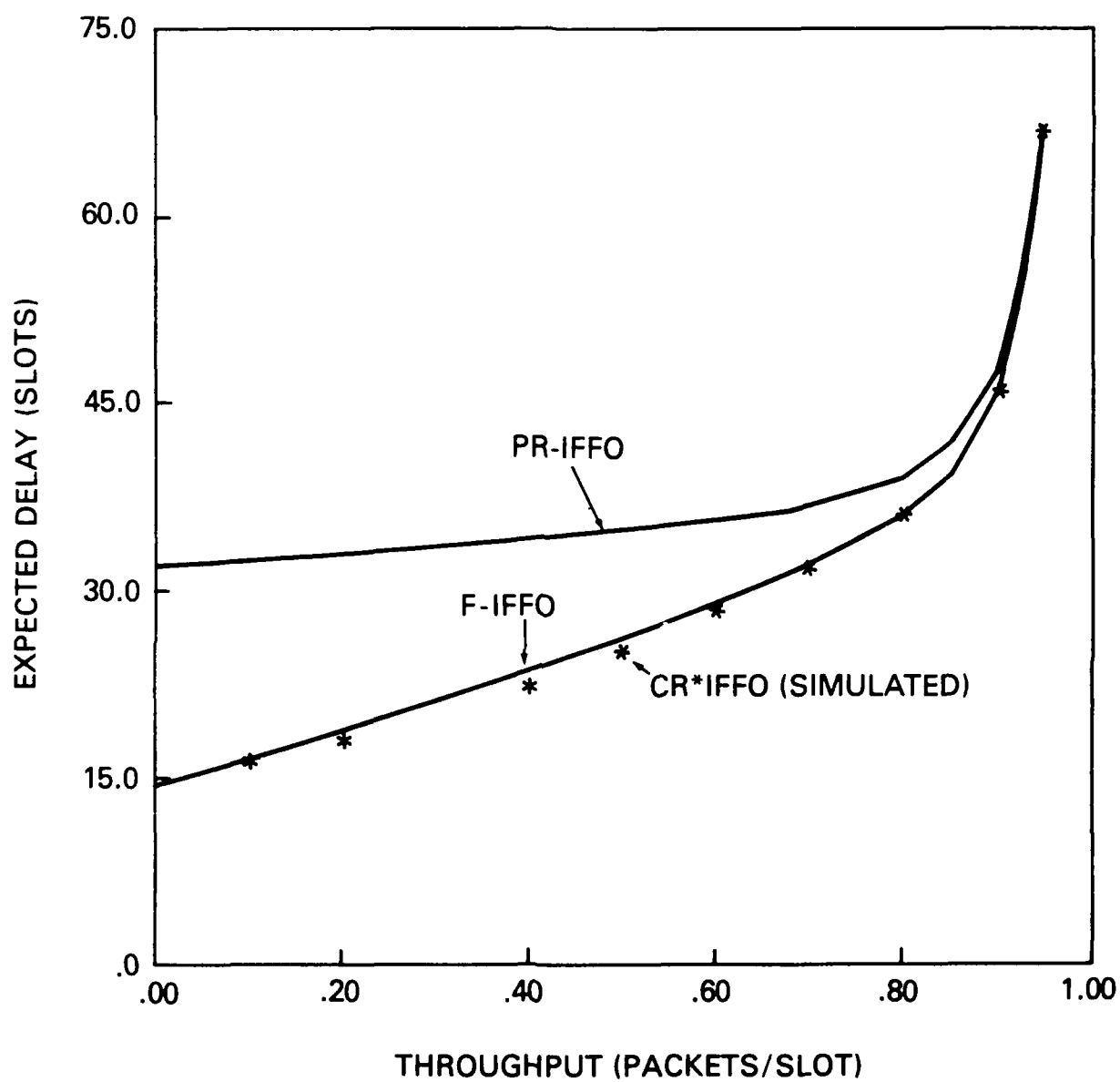Fig. 3   Operation of the IFFO protocols, showing transmission procedure for frame-k arrivals.

$R_{k+2} = A_k - S_k$ RESERVED SLOTS IN FRAME $k + 2$

S_k SUCCESSFUL CONTENTION TRANSMISSIONS

STATUS SLOT

R_k N_k R_{k+1} N_{k+1} R_{k+2} N_{k+2}

FRAME $k + 1$ FRAME $k + 2$

A_k ARRIVALS IN FRAME k

Fig. 4   Delay-throughput curves for PR-IFFO, F-IFFO, and CR*IFFO
($R = 12$, $M = 10$)
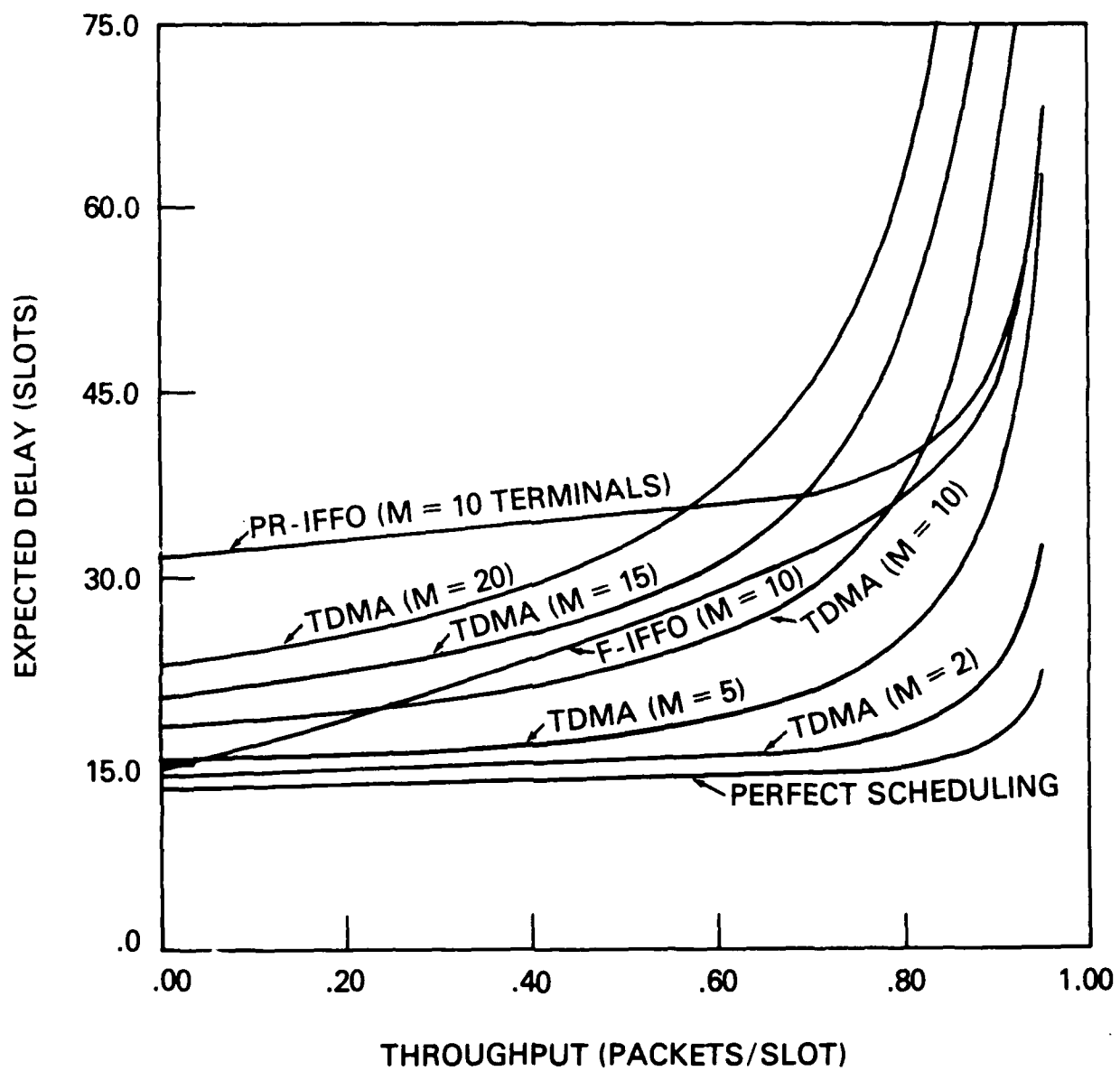
Fig. 5   Delay-throughput curves for PR-IFFO, F-IFFO, TDMA, and
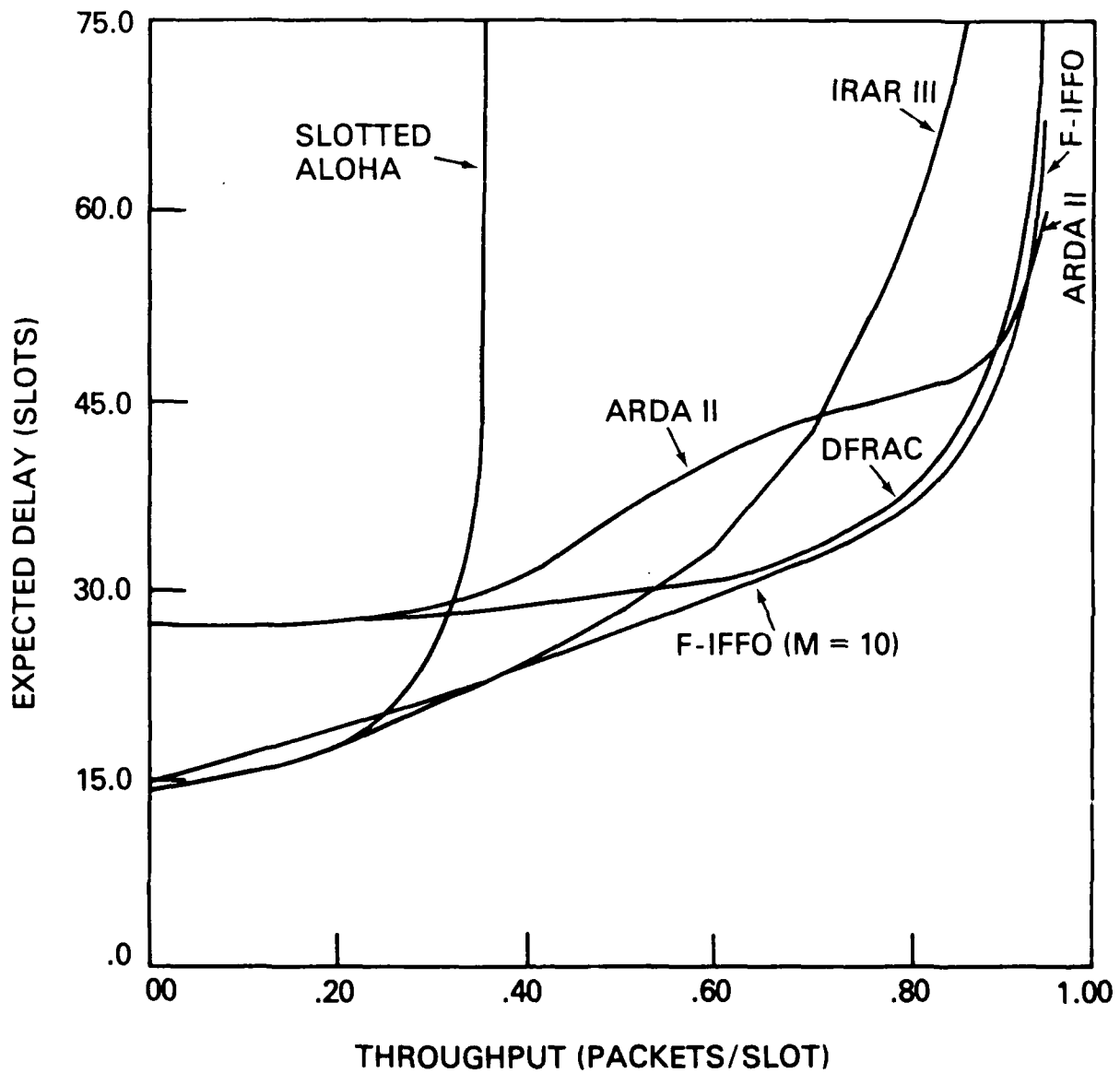"perfect scheduling" ($R = 12$).

Fig. 6   Delay-throughput curves for F-IFFO, ARDA II, IRAR III,
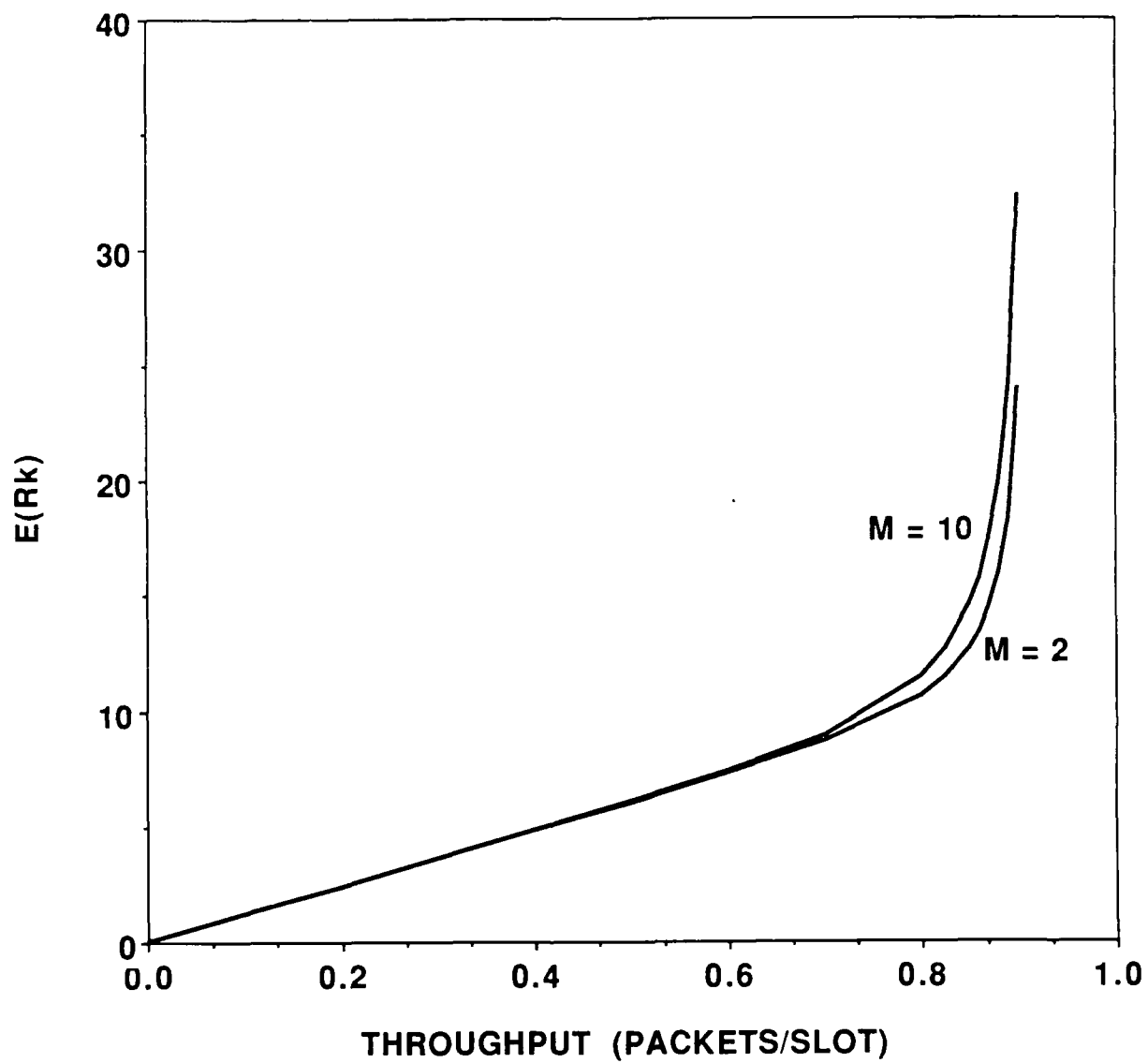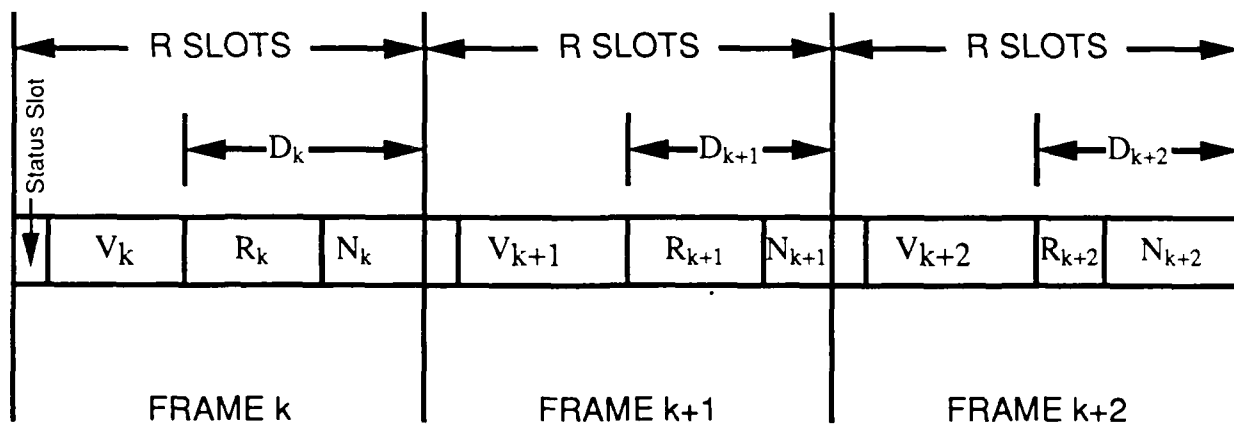DFRAC, and slotted ALOHA ($R = 12$).

Fig. 7 Expected value of R_k under PR-IFFL (R = 12, M = 2, 10)

Fig. 8   Frame structure for the VD-IFFL protocols